

## **A Complex Case-based Advisor**

M. Lourdes Borrajo

Department of Computer Science, University of Vigo, Campus As Lagoas, s/n, 32004 Ourense, Spain. E-mail: [lborrajo@uvigo.es](mailto:lborrajo@uvigo.es)

Rosalía Laza

Department of Computer Science, University of Vigo, Campus As Lagoas, s/n, 32004 Ourense, Spain. E-mail: [rlaza@uvigo.es](mailto:rlaza@uvigo.es)

Juan M. Corchado

Departamento de Informática y Automática,  
University of Salamanca, Plaza de la Merced s/n, 37008 Salamanca, Spain. E-mail: [corchado@usal.es](mailto:corchado@usal.es)

*The complexity of current organization systems, and the increase in the importance of the realization of internal controls in firms, make it necessary to construct models that automate and facilitate the work of auditors. A tool for the decision support process has been developed. The objective of the system is to facilitate the process of internal auditing in small and medium-sized enterprises. The system analyses the data that characterises each one of the activities carried out by the firm, then determines the state of each activity, calculates the associated risk, detects the erroneous processes and generates recommendations to improve these processes. The developed model is composed of two case-based reasoning systems. One is used to identify the activities that may be improved and the other to determine how the activities could be improved. Each of the two subsystems uses a different problem solving method in each of the steps of the reasoning cycle. The system has been tested in 22 small and medium-sized companies in the textile sector, located in the northwest of Spain during the last 38 months and the obtained results are very encouraging.*

## **1. Introduction**

Nowadays, organization systems employed in enterprises are increasing in complexity. Moreover, in recent years, the number of regulatory norms has increased considerably. As a consequence of this, the need for periodic internal audits has arisen. However, the evaluation and the prediction of the evolution of these types of systems, characterized by their great dynamism, are, in general, complicated. It is necessary to construct models that facilitate the analysis of work carried out in changing environments, such as finance.

The processes carried out inside a firm are grouped in functional areas (Corchado et al., 2004) denominated “Functions”. A Function is a group of coordinated and related activities, which are necessary to reach the objectives of the firm and are carried out in a systematic and iterative way (Mas and Ramió, 1997). The functions that are usually carried out within a firm are: Purchases, Cash Management, Sales, Information Technology, Fixed Assets Management, Compliance to Legal Norms and Human Resources. In turn, each one of these functions is broken down into a series of activities (see Figure 1). For example, the function Information Technology is divided in the following activities: Computer Plan Development, Study of Systems, Installation of Systems, Treatment of Information Flows and Security Management.

Each activity is composed of a number of tasks. Control procedures also have to be established in the tasks to ensure that the established objectives are achieved.

<Insert Figure 1 here>

The developed system is composed of two fundamental subsystems (Borrajo, 2003; Borrajo et al., 2005):

- ISA Subsystem (Identification of the State of the Activity) whose objectives are: to identify the state or situation of each one of activities of the company and to calculate the risk associated with this state.
- GR Subsystem (Generation of Recommendations), whose goal is to generate recommendations to reduce the number of inconsistent processes in the firm.

Both subsystems are implemented using a case-based reasoning (CBR) system (Kolodner, 1993; Aamodt and Plaza, 1994; Watson, 1997; Lenz et al., 1998). The CBR system associated with each subsystem uses different problem solving techniques in every phase and both subsystems share the same case memory (Hunt and Miles, 1994; Medsker, 1995).

The first subsystem uses a GCS network (*Growing Cell Structures*) (Köhle and Merkl, 1996) in the retrieval phase whereas the ANFIS (*Adaptive Neuro-Fuzzy Inference System*) (Nauck, 1997) model implements the adaptation phase. The second subsystem uses a different problem solving method in each of the steps of the reasoning cycle: fuzzy clustering during the retrieval phase, a multi-criterion discreet method during the reuse phase and a rule based system for recommendation generation. The main advantage of the proposed solution is that, once implemented, the CBR system gains in autonomy, accuracy and provides the users with a higher level of confidence in the system.

The rest of this article is structured as follows: first, the concept of internal control (IC) is explained and its relevance in the present business world is described. Secondly, the advantages of using the methodological framework that provide the CBR systems for the internal control problem are discussed. Then, the proposed CBR based model is analysed and its results are evaluated. Finally, the conclusions are presented.

## **2. Internal Control**

Small to medium enterprises require an internal control mechanism in order to monitor their *modus operandi* and to analyse whether they are achieving their goals. Such mechanisms are constructed around a series of organizational policies and specific procedures dedicated to giving reasonable guarantees to their executive bodies. This group of policies and procedures are named "controls", and they all conform to the company's internal control structure. The establishment of objectives (which is not a function of the Internal Control) is a previous condition for control risk evaluation, which is the main goal of the Internal Control.

In general terms, the administration of a firm has three large categories of objectives when designing an internal control structure (AICPA, 1996):

1. Reliability of financial information.
2. Efficiency and effectiveness of operations.
3. Fulfillment of the applicable rules and regulations.

The internal auditor must monitor the internal controls directly and recommend improvements on them. Therefore, all the activities carried out inside the organization can be included, potentially, within the internal auditors' remit. Essentially, the activities of the auditor related to IC can be summarized as follows:

- To be familiar with and possess the appropriate documentation related to the different components of the system that could affect financial aspects.
- To assess the quality of internal controls in order to facilitate the planning of the audit process with the aim of obtaining necessary indicators.
- To assess internal controls in order to estimate the level of error and reach a decision on the final opinion to be issued in the memorandum on the system under consideration.

As a consequence of the great changes in firms brought, about by current technological advances, considerable modifications have taken place in the area of auditing, basically characterized by the following features (Sanchez, 1995):

- Progressive increase in the number and level of complexity of audit rules and procedures.
- Changes in the norms of professional ethics, which demand greater control and quality in auditing.
- Greater competitiveness between auditing firms, consequently resulting in lower fees; the offer of new services to clients (e.g. financial or computing assessment ...).

- Development of new types of auditing (e.g. operative management auditing, computer auditing, environmental auditing ...).

Together, these circumstances have made the audit profession increasingly competitive. Consequently, the need has arisen for new techniques and tools, which can be provided by information technology and artificial intelligence. The aim is to achieve more relevant, more suitable information, in order to help auditors make decisions faster and thereby increase the efficiency and quality of auditing. The following section presents a system developed to facilitate the auditing process.

### **3. Case-Based Reasoning for Internal Control**

Rule-based systems (RBS) have traditionally been used with the purpose of delimiting the audit decision-making tasks (Denna et al., 1991). However, Messier and Hansen (1988) found many situations in which auditors resolved problems by referring to previous situations. This contrasts with the very nature of RBS systems, since they have very little capacity for extracting information from past experience and present problems in order to adapt to changes in the environment.

In contrast, case based reasoning (CBR) systems are able to relate past experiences or cases to current observations, solving new problems through the memorization and adaptation of previously tested solutions. This is an effective way of learning, similar to the general structure of human thought. CBR systems are especially suitable when the rules that define a knowledge system are difficult to obtain, or the number and complexity of the rules is too large to create an expert system. Moreover, CBR systems have the capacity to update their memory dynamically, based on new information (new cases), as well as, improving the resolution of problems (Riesbeck and Schank, 1989).

However, in problems like those presented in this study, standard monitoring and prediction techniques cannot be applied due to the complexity of the problem, the existence of certain preliminary knowledge, the great dynamism of the business world, and so forth. In these types of systems it is necessary to use models that combine the advantages of several techniques capable of solving different aspects of the whole problem and of incorporating updated knowledge.

In this sense, a CBR-based adaptive system has been developed. The system possesses the flexibility to behave in different ways and to evolve, depending on the environment in which it operates. A case based reasoning system solves a given problem by means of the adaptation of previous solutions to similar problems (Aamodt and Plaza, 1994). The CBR memory stores a certain number of cases. A case includes a problem and the solution to this problem. The solution of a new problem is obtained recovering similar cases stored in the CBR memory.

A CBR is a dynamic system in which new problems are added continuously to its memory, the similar problems are eliminated and gradually new ones are created by combining several existent ones. This methodology is based on the fact that humans use the knowledge learned in previous experiences to solve present problems.

CBR systems record past problem solving experiences and, by means of indexing algorithms, retrieve previously stored problems with their solutions (cases), and match and adapt them to a given situation. This means that the set of cases stored in the memory of CBR systems represents the knowledge concerning the domain of the CBR. As discussed below, this knowledge is updated constantly.

A typical CBR system is composed of four sequential steps which are recalled every time a problem needs to be solved (Kolodner, 1993; Aamodt and Plaza, 1994; Watson, 1997):

1. *Retrieve* the most relevant case(s).
2. *Reuse* the case(s) in order to solve the problem.
3. *Revise* the proposed solution if necessary.
4. *Retain* the new solution as a part of a new case.

Like other mechanisms of problem solving, the objective of a CBR is to find the solution for a certain problem. A CBR is a system of incremental learning, because each time a problem is solved, a new experience is retained, thereby making it available for future reuse.

CBR systems have proven to be an effective method for problem solving in multiple domains, for example, prediction, diagnosis, control and planning (Lenz et al., 1998). This technology has been successfully used in several disciplines: law, medicine, diagnosis systems and so on (Watson, 1997).

The case based reasoning can be used by itself, or as part of another conventional or intelligent system (Medsker, 1995). Although there are many successful applications based on CBR methods alone, CBR systems can be improved by combining them with other technologies (Hunt and Miles, 1994). Their suitability for integration with other technologies, creating a global hybrid reasoning system, stems from the fact that CBR systems are very flexible algorithms, capable of absorbing the beneficial properties of other technologies.

#### **4. Neuro-symbolic System for Internal Control**

This section describes the internal control system in detail. Although the aim is to develop a generic model useful in any type of small to medium enterprise, the initial work has focused on the textile sector to facilitate the research and its evaluation. The model presented here may be extended or adapted to other sectors. Twenty two



companies from the North-west of Spain have collaborated in this research, some of them manufacture products for major firms such as Inditex and others work mainly for the Spanish market. The companies have different levels of automation and all of them were very interested in a tool such as the one developed in the framework of this investigation. After analysing the data relative to the activities developed within a given firm, the constructed system is able to determine the state of each of the activities and calculate the associated risk. It also detects any erroneous processes and generates recommendations for improving these processes. As shown below the problem solving mechanism developed makes its decision with the help of a couple of CBR systems whose memory has been fed with cases constructed with information provided by the firm and with prototypical cases identified by 34 internal control experts who have collaborated and supervised the developed model.

The cycle of operations of the developed case based reasoning system is based on the classic life cycle of a CBR system (Aamodt and Plaza, 1994; Watson and Marir, 1994). This cycle is executed twice, since the system bases its operation on two CBR subsystems (the ISA subsystem -Identification of the State of the Activity and the GR subsystem -Generation of Recommendations), as can be seen in Figure 2. Both subsystems share the same case base and a case represents the “shape” of a given activity developed in the company.

<Insert Figure 2 here>

Every time that it is necessary to obtain a new estimate of the state of an activity, the hybrid system progresses through several phases. This evolution allows the system, on the one hand, (i) to identify the latest situations most similar to the current situation, (ii) to adapt the current knowledge to generate an initial estimate of the risk of the activity being analysed, and on the other hand, (iii) to identify old situations that serve as a basis

to detect the erroneous processes developed within the activity and (iv) to select the best of all possible activities. The activity selected will then serve as guide for the establishment of a set of recommendations that allow the activity, its function and the company itself to develop in a more positive way. The retention phase guarantees that the system evolves in parallel with the firm, basing the corrective actions on the calculation of the error previously made. The following sections describe the different phases of the proposed model.

#### **4.1. Data Acquisition**

The data used to construct the model were obtained from a set of surveys:

1. External auditors' surveys. Through the results of the surveys, each one of the functions and activities of a firm is assigned a level importance. This information allows the system to calculate the control risk associated with an activity. It also allows the system to prioritise the recommendations for improving the business processes. This data is stored in the system database.
2. Experts' surveys on the different functional areas. The second type of survey was carried out among several experts in the different functional areas of various firms. This type of survey attempts to reflect the experience of the experts in their different fields. For each activity, the survey presents two possible situations: the first one tries to reflect the situation of an activity with an incorrect activity state and the second one tries to reflect the situation of an activity with a satisfactory activity state. Both situations will be valued by the expert using a percentage. The data acquired by the surveys has been used to build the prototype cases for the initial case base.

Table 1 shows the case structure that constitutes the case base.

<Insert Table 1 here>

Each case is composed of the following attributes:

- *Case number*: Unique identification: positive integer number.
- *Input vector*: Information about the tasks (n sub-vectors) that constitute an industrial activity:  $((IR_1, V_1), (IR_2, V_2), \dots, (IR_n, V_n))$  for n tasks. Each sub-vector task has the following structure  $(GI_i, V_i)$ :
  - $IR_i$ : importance rate for this task within the activity. It can only take one of the following values: VHI (Very high importance), HI (High Importance), AI (Average Importance), LI (Low Importance), VLI (Very low importance)
  - $V_i$ : Value of the realization state of a given task: a positive integer number (between 1 and 10).
- *Function number*: Unique identification number for each function
- *Activity number*: Unique identification number for each activity
- *Reliability*: Percentage of probability of success. It represents the percentage of success obtained using the case as a reference to generate recommendations.
- *Degree of membership*:  $((n_1, \mu_1), (n_2, \mu_2), \dots, (n_k, \mu_k))$ 
  - $n_i$ : represents the  $i^{st}$  cluster
  - $\mu_i$ : represents the membership value of the case to the cluster  $n_i$
- *Activity State*: degree of perfection of the development of the activity, expressed by a percentage.

#### 4.2. ISA Subsystem (Identification of the State of the Activity)

The ISA subsystem identifies the state or situation of each of the firm's activities and calculates the risk associated with this situation (see Fig. 3). The system uses the data for the activity, introduced by the firm's internal auditor, to construct the problem case. For each task that makes up the analysed activity, the problem case is composed of the

value of the realization state for that task, and its level of importance within the activity (according to the internal auditor).

In this way, a problem case for an activity of  $n$  tasks, will be composed of a vector such as:  $((IR_1, V_1), (IR_2, V_2), \dots, (IR_n, V_n))$  where:

- $IR_i$ : importance rate for this task within the activity. It can only take one of the following values: VHI (Very high importance), HI (High Importance), AI (Average Importance), LI (Low Importance), VLI (Very low importance).
- $V_i$ : Value of the realization state of a given task. It is a positive integer number (between 1 and 10).

<Insert Figure 3 here>

#### **4.2.1. Retrieval Phase**

The retrieval stage must be carried out using a method that guarantees the retrieval of a reasonably small number of cases that are related to the current problem. We have experimented with a number of different methods for evaluating the similarity of the current problem with those in the cases base: K-means, Sparse Kernel Principal, Principal Component Analysis, the Self Organising Map and the Scale Invariant Map. The best results have been obtained with the Crowing Cell Structures (GCS) method.

The type of GCS used in this work is characterized by a two-dimensional space, where the units (cells) are connected and organised in triangles. Each cell in the network is associated with a weight vector,  $w$ , which has the same dimension as the input data. At the beginning of the learning process, the weight vector of each cell is initialised with random values. The basic learning process in a GCS network consists of topology modification and weight vector adaptations (Köhle, 1996). This vector is the prototype case of each network cell.

For each training case, the network performs a so-called learning cycle, which may result in topology modification and weight vector adaptation. In the first step of each learning cycle, the cell  $c$ , with the smallest distance between its weight vector,  $w_c$ , and the actual input vector,  $x$ , is chosen as the winner cell or best-match cell (see Eq. (1)).

$$c : \|x - w_c\| \leq \|x - w_i\|; \forall i \in O \quad (1)$$

The second step consists of the adaptation of the weight vectors of the winning cell and their neighbouring cells; see Eqs. (2) and (3). The terms  $\varepsilon_c$  and  $\varepsilon_n$  represent the learning rates for the winner and its neighbours respectively. Both learning rates are constant during learning, and  $\varepsilon_c, \varepsilon_n \in [0, 1]$ .

$$w_c(t+1) = w_c(t) + \varepsilon_c(x - w_c) \quad (2)$$

$$w_n(t+1) = w_n(t) + \varepsilon_n(x - w_n), \forall n \in N_c \quad (3)$$

In the third step of a learning cycle, each cell is assigned a *signal counter*,  $\tau$ , that reflects how often a cell has been chosen as the winner (see Eqs. (4) and (5)).

$$\tau_c(t+1) = \tau_c(t) + 1 \quad (4)$$

$$\tau_i(t+1) = \tau_i(t) - \alpha\tau_i(t); i \neq c \quad (5)$$

The parameter  $\alpha$  reflects a constant *rate of counter reduction* for the rest of the cells in the current learning cycle. Growing cell structures also modify the overall network structure by inserting new cells into those regions that represent large portions of the input data. The frequency of insertion update is controlled by the parameter  $\lambda$ , which is associated with the number of learning cycles between two cell insertions (see Eqs. (6), (7) and (8)).

$$h_i = \frac{\tau_i}{\sum_j \tau_j}; \forall i, j \notin O \quad (6)$$

$$q : h_q \geq h_i; \forall i \in O \quad (7)$$

$$r : \|w_r - w_q\| \geq \|w_p - w_q\|; \forall p \in N_q \quad (8)$$

The GCS network indicates the prototype case of each node, its topology (Fdez-Riverola, 2002) and calculates the scale parameters  $\sigma_j$  of each node. This parameter measures the width in the gauss membership function. It can be seen in Figure 4. Higher values of  $\sigma$  provide an area of the dominant node in the environment of the centroid more extended.

<Insert Figure 4 here>

To calculate the node  $j$ , the prototype cases of its neighbour nodes are selected; then the average of the square-distance between them is calculated (Warwick, 1995; Wasserman, 1993), that is:

$$\sigma_{jk}^2 = \frac{1}{N} \sum_{l=1}^N \|c_l - c_j\|^2 = \frac{1}{N} \sum_{l=1}^N (c_{lk} - c_{jk})^2$$

This information is used for making the initial fuzzy inference system, which is used by the ANFIS model. This fuzzy inference system has a set of TSK rules; each node provides a fuzzy rule. The rules have the form:

$$R_j : \text{if } x_1 \text{ is } A_{1j} \text{ and } x_2 \text{ is } A_{2j} \text{ and } \dots \text{ and } x_M \text{ is } A_{Mj}, \text{ then } y = g_j(x_1, x_2, \dots, x_M)$$

where  $g_j(\cdot)$  is a polynomic function in  $x_i$ .

Each attribute is represented by a gauss function (Eq. (9)), which makes part of the antecedent of the rule.

$$A_{ij}(x_i) = \exp\left(-\frac{(c_k - x_i)^2}{2\sigma_k^2}\right) \quad (9)$$

where  $c$  is a prototype case and  $\sigma$  the distance.

The next step consists of obtaining the consequents of each TSK rule. The method applied is least-square (Nauck, 1997). This initial fuzzy inference system is adapted with the ANFIS model.

The problem is determined by a set of variables with particular values, which are used as inputs of the GCS net. Next, the search for the node, which the new problem belongs to, is started, that is, the winner node must be found. This node is obtained by calculating the Euclidean distance between the new problem and every prototype case of each case. The node with the least distance will be the winner. All the cases associated to the winner node will be considered the most similar ones, and will be used in the ANFIS model.

#### **4.2.1.1 Advantages**

GCS neural networks are a variation of the Kohonen's Self-Organising Maps and provide the basis for powerful information retrieval applications and similarity visualization tools. They offer several advantages, such as:

1. The network structure is automatically determined from the domain data and thus reflects the underlying characteristics of the domain in the very topology of the network.
2. In contrast to Kohonen self-organising maps, the GCS self-organising model consists of a small number of constant parameters. Therefore, there is no need to define time-dependent or schedule parameters.

3. In contrast to Kohonen self-organising maps, the adaptation rules of the GCS also enable the processing of data with changing probability distributions.
4. Its ability to perform problem-dependent error measures allows the implementation of better adaptive data representations (insertion and deletion of cells) in comparison with static-topology models.
5. Its facility to insert and delete cells allows the accurate estimation of probability densities of input signals.
6. Its ability to interrupt a learning process or to continue a previously interrupted one permits the construction of incremental and dynamic learning systems.
7. GCS networks have demonstrated their capacity to process both small and high dimensionality data in several application domains in contrast to back propagation networks for instance.

#### **4.2.2. Reuse Phase**

This phase aims to obtain an initial estimation of the state of the analysed activity. In order to obtain this estimation, the ANFIS network (*Adaptive-Network-based Fuzzy Inference Systems*) is used.

One of the first hybrid neuro-fuzzy systems for function approximation was Jang's ANFIS model (Nauck, 1997). ANFIS adjusts only the membership functions of the antecedents and the consequent parameters.

Due to ANFIS only using differentiable functions, it is easy to apply standard learning procedures from neural network theory. For ANFIS a mixture of back propagation (gradient descent) and *Least Squares Estimation* (LSE) is used. Back propagation is used to discover the antecedent parameters, i.e. the membership functions, and LSE is used to determine the coefficients of the linear combinations in the rule's consequents.



A step in the learning procedure has two parts, which are shown in Table 2. In the first part the input patterns are propagated, and the optimal consequent parameters are estimated by an iterative least mean squares procedure, while the antecedent parameters are assumed to be fixed for the current cycle through the training set. In the second part the patterns are propagated again, and in this, epoch back propagation is used to modify the antecedent parameters, while the consequent parameters remain fixed.

<Insert Table 2 here>

The cases retrieved in the previous one are used, that is to say, the inference system provided by the GCS net. While the result is a fuzzy inference system adapted to solve a particular problem.

#### **4.2.2.1 Advantages**

ANFIS is one of the best tradeoffs between neural and fuzzy systems, providing *smoothness*, due to the Fuzzy Control interpolation and *adaptability*, due to the neural network back propagation. Following (o subsequently) are the advantages of the ANFIS model.

1. Translates prior knowledge into network topology and initial fuzzy partitions.  
There is a faster convergence than typical feed forward neural network.
2. Uses data to determine TSK rules. It is a model compactness (smaller rules than using labels).
3. Uses *Least Squares* algorithm to compute coefficient polynomials. Uses *back propagation* to tune fuzzy partitions. The advantage is the automatic TSK parametric tuning.

Uses inference mechanism to interpolate among rules. The advantage is smoothness guaranteed by interpolation.

#### **4.2.3. Revision Phase**

The objective of this phase is to guarantee or refute the initial solution proposed by the ANFIS network, obtaining a final solution and working out, apart from this, the control risk.

Faced with the initial estimation generated by the ANFIS network, the internal auditor will be responsible for deciding whether to accept the aforementioned solution. It will be based on the specific knowledge it possesses about the company in which it is working. If the estimation provided by the system is considered to be valid, the system will take this solution as the final solution and, in the next phase of the CBR cycle, will store a new case in the case base, formed from the problem case and the final solution.

The system will assign this new case with an initial reliability of 100%. If, on the contrary, the internal auditor considers that the solution provided by the system is not valid, they will provide their own solution that the system will take as the final solution. This, along with the problem case, will be stored as a new case in the case base in the following phase. This new case will be assigned a reliability of 30%. This value has been determined bearing in mind the opinion of various auditors as to the importance that should be assigned to the personal opinion of the internal auditor.

As of the final solution: the state of the activity, the system must work out the control risk associated with the activity. All developed activity in the business sector has an associated risk that will indicate the negative influence that it exerts over the proper running of the business. That is to say, the control risk of an activity measures the impact, on the whole on all of the managerial processes that the current state of the aforementioned activity represents. In this investigation it is considered that the risk level can assume one of these three possible values: low, medium and high. The calculation of the control risk level associated with an activity is based on the current

state of the activity and on its level of importance. This last value is obtained after analysing the information obtained from a series of surveys (98 in total) answered by auditors all over Spain. In the previously mentioned surveys the auditors were asked to rank from 1 to 10 the importance that they gave the above-mentioned activity together with the function to which it belongs. The greater the importance is, the greater weighting the activity will have within the internal control system.

Then, as of the level of importance of the activity, and bearing in mind the final solution obtained after the revision phase, the control risk level can be worked out. The calculation of the control risk level is carried out using if-then rules in which the importance that the auditors have assigned that activity is compared to the final solution or state of the activity. These rules follow the pattern shown in Figure 5, where “auditors\_importance” indicates the average level of importance assigned to the previously mentioned activity by the external auditors and “activity\_state” is the final solution or state of the activity.

<Insert Figure 5 here>

Figure 6 shows the pseudocode for this revision stage where  $V_p$  represents the vector of attributes that describe a new problem (problem case),  $m_a$  is the mean importance assigned to this activity by the auditors,  $s_i$  is the initial estimation carried out by the ANFIS network,  $s_f$  is the generated final solution and  $n_r$  represents the obtained risk level.

<Insert Figure 6 here>

#### **4.2.4 Retain Phase**

The last stage of the life cycle of the Identification of the Activity State subsystem is incorporating what has been learnt into the system’s memory after solving the new problem. Once the revision phase has been surpassed, after obtaining the final solution,

a new case is built (*problem + solution*), which is stored in the system's memory. Apart from the global update of knowledge that the insertion of a new case into the system's memory entails, the presented hybrid system carries out a local adaptation of the knowledge structures that it makes use of.

The fuzzy cluster system, which contains the prototypes regarding the activity to which the new case corresponds, is reorganised to answer the appearance of this new case, modifying its internal structure and adapting to the available new knowledge.

In the same way, the GCS network uses the new case to carry out a complete learning cycle, updating the position of its centres and modifying the value of its weightings that connect the intermediate layer to the exit layer.

Learning is a continuous process, in which the ANFIS network acts as a mechanism capable of absorbing the knowledge of a certain number of cases, and generalising starting from them. During this process, the ANFIS network interpolates and makes predictions without forgetting part of the ones that have already been previously carried out. The system's memory acts as a permanent memory capable of keeping a lot of cases or experiences, while the ANFIS network acts a short term memory, capable of recognising recently learnt patterns and generalising starting from them.

Table 3 shows a summary of the technologies used in each stage of the Identification of the State of the Activity subsystem. For each one of them the input details, the process carried out and the output details that will be used in the subsequent stages of the estimation process are shown.

<Insert Table 3 here>

### 4.3 GR Subsystem (Generation of Recommendations)

The objective of this subsystem is, once the stages of the previous subsystem have concluded, to carry out the recommendations that will help the internal auditor to decide on the actions to take with the aim of improving the business's external and internal processes (see figure 7). This subsystem is totally dependent on the previous one, given that part of the case (problem + solution) is generated in the ISA-Identification Subsystem of the State of the Activity (see Figure 2).

<Insert Figure 7 here>

#### 4.3.1 Retrieval phase

The function of the second subsystem is to generate the recommendations that guide the internal auditor in the task of deciding on the actions to take in order to improve the state of the analysed activity. In order to be able to recommend changes in the execution of the managerial processes, it is necessary to compare the current situation of the activity, represented by *problem case + solution*, generated in the State of the Activity subsystem, with those cases from the case base that reflect a better managerial administration.

With this aim, only the cases that are most similar to the problem case will be worked with. Given that in the retrieval stage of the Identification of the Activity State subsystem the cluster whose centre was closest to the problem case was identified, this same cluster will be used to recuperate the cases that will be used in the next adaptation phase. The subsequent process in this retrieval stage is based on the use of query relaxation (Gardingen and Watson, 1998). As a result, those cases that fulfil the following conditions are initially recuperated from the case base:

1. Its solution or activity state must be superior, to the one that was generated as a final solution in the previous subsystem, by an interval between 15% and 20%.  
If enough cases are not retrieved (25 is considered to be enough) the percentages are relaxed further, increasing the percentage interval range by 5%.
2. Furthermore, they must possess a level of pertinence to the higher cluster of 75% and a level of reliability greater than 50%. These two constant values have been established by the auditors who participated in the survey.

Figure 8 shows the diagram of the adopted retrieval process, where  $X$  stands for the case group which represents the knowledge of a determined activity that exists within the memory of the system,  $v_p$  represents the vector of attributes that describes the problem case,  $s_f$  is the final solution generated in the ISA subsystem as a solution to the problem case,  $s_i$  is the solution to case  $i$ ,  $m$  is the identifier of the cluster whose centre is closest to the problem case,  $U$  is the matrix of pertinences,  $u_{mi}$  is the level of pertinence of case  $i$  to cluster  $m$  and  $K$  is the most relevant case group retrieved.

<Insert Figure 8 here>

#### 4.3.2 Reuse Phase

Given that the objective of this subsystem is to generate a series of recommendations from the problem case, it is necessary to search for a case from the case base (or a combination of various cases) which serves as a guide to generate recommendations, comparing this case with the problem case. This comparison will allow the company to detect which processes must be modified – in other words, which tasks must be improved.

As already explained above, in the retrieval phase, the cases obtained are those which best reflect the most favourable overall activity state presented by the analysed activity. From all these cases, in this adaptation phase, the subsystem will have to select the case

which maximises the value of each of the tasks ( $V_i$ ), taking into account the degree of importance ( $IR_i$ ) or weighting that each task has for the overall activity.

In this way, the problem of selecting a case from all those retrieved can be made similar to a multi-criteria decision-making problem where the alternatives are the different retrieved cases and the objective is to maximise the task values (which will represent the attributes).

In this study, the initial version of the Electre method has been selected (Barba-Romero and Pomeral, 1997; Romero, 1993) in order to tackle the problem of choosing one of the alternatives. The Electre method proposes a strategy for reducing the size of all the possible solutions, through the segregation of a group of the most favourable cases opposite another group which includes the least favourable cases. The application of the aforementioned method will produce the selection of one or various cases from among those retrieved.

The Electre method is based on the fact that the preferential weightings vector, which is subjectively associated with each attribute, is known. As in this study, the weighting of an attribute (represented by its level of importance) is different for each alternative, it is necessary to obtain a unique weightings vector for the attributes of the group of alternatives or retrieved cases. In this case, the obtaining of the weighting vector is carried out by calculating, for each attribute, the mean weightings of the attribute in question, for all the alternatives.

On the other hand, Electre returns, as a solution, the best alternative, or group of best alternatives in the event of there not being any that prevails over the rest. Given that for the generation of recommendations it is necessary to begin with a unique alternative, it has been considered that if the output of the multi-criteria decision method are various

alternatives, the combination of these alternatives will be adopted as the output, obtained starting from the mean of the values for each attribute.

Figure 9 shows the pseudocode for this adaptation phase where  $K$  is the group of most relevant cases retrieved in the previous phase,  $v_{el}$  is the case or alternative obtained after the adaptation phase starting from the group of cases  $K$ ,  $v_{el(j)}$  is the value of the attribute  $j$  of the case  $v_{el}$ ,  $C$  is the group of alternatives or cases obtained as output of the Electre method,  $C_{ij}$  is the value of the attribute  $j$  belonging to the alternative  $i$  and  $n$  is the number of alternatives generated by the Electre method.

<Insert Figure 9 here>

The case obtained as a result of applying the Electre method represents the objective to be reached for the analysed activity or the standard to be followed in order to fulfil the objectives of the company or, specifically, the objective associated with the activity.

In this way, the function of the recommendations which are subsequently generated will be to make sure that the different tasks that form the problem case reach a situation that is as similar as possible to the case obtained as output of the Electre method.

In this way, in order to generate the recommendations, the output from the Electre method is compared to the problem case, comparing the values ( $V_i$ ) of each of the attributes or tasks in both cases. The objective is to detect which tasks must be improved, establishing an order of priorities in terms of the weighting ( $IR_i$ ) of each task over the activity as a whole. In other words, to identify the possible deviations of the activity and to evaluate the extent of the deviations in terms of the tasks' level of importance ( $IR_i$ ). In this way, the system generates recommendations related to the inconsistent processes found, that is, the differences between the attribute values in the problem case and the attributes in the objective case (considered the standard) obtained with the Electre method, representing the auditor's potential recommendations.



The group of attributes of stored cases in the case base represent the group of values that as much the experts in each activity as the auditors have judged to be effective (in the surveys carried out) in terms of the characteristics of the company. Given that the characteristics of the current case (problem) are similar to the obtained objective case, the auditor can argue that the attribute values must also be similar. This provokes a more convincing argument than the one of basing it on probabilities and estimated losses or risks.

The generation of control recommendations by comparing the values of the current case with those of past cases also eliminates other problems such as the lack of outputs or pre-defined results. Many possible values exist along with a large number of combinations that could be included in the auditor's recommendations. However, not all of the combinations are valid; some combinations may not be feasible or make sense.

In contrast to the CBRs, as much the expert systems as the neuron networks will need to have possible outputs specified for them previously.

Based on the predictions and recommendations generated by the system, the internal auditor can inform the company of inconsistent processes and the measures to be taken. This is a decision support system that facilitates the auditing process for internal auditors.

#### **4.3.3 Retain Phase**

After the time necessary for correcting the detected errors, the firm is evaluated again. Auditing experts consider that three months is long enough for the company to evolve towards a more favourable state. If it is verified that the erroneous processes and the level of risk have diminished, the retention phase is carried out, modifying the case used to generate the recommendations. The reliability (percentage of successful identifications obtained with this case) of this case is thereby increased by 10%. In

contrast, when the firm has not evolved to a better state, the reliability of the case is decreased by 10%. Furthermore, those cases whose level of reliability is less than 15% are eliminated, regrouping the remaining cases in fuzzy clusters once more.

Table 4 shows a summary of the technologies used in each stage of the Generation of Recommendations subsystem. For each one of them the input details, the process carried out and the output details are shown.

<Insert Table 4 here>

#### 4.4. Case Study

Following an example to facilitate the comprehension of the system's way of working is put forward. Let us assume that the company X's internal auditor introduces the information that reflects the "Selection of Suppliers" activity's current situation into the system, belonging to the "Acquisition of Goods and Services" operative area. This activity possesses seven tasks. For each task its level of importance ( $IR_i$ ) within the activity and its level of execution ( $V_i$ ) must be specified. Table 5 shows an example.

<Insert Table 5 here>

So that numerical operations with the level of importance attribute can be carried out, it will be assessed by assigning a numerical value to each of the possible options.

<Insert Table 6 here>

In this way, the problem case will have the following aspect:

$((2,1), (4,8), (4,8), (5,9), (4,6), (2,2), (4,5))$

In view of this new problem, the first subsystem, the ISA (Identification of the State of Activity) subsystem, in its retrieval phase, compares the problem case with each one of the fuzzy cluster centres, looking for the most similar.

Let us assume that the case base cases relating to the “Selection of Suppliers” activity are grouped into three fuzzy clusters. A centre or representative of the cluster exists for each one of them. If we assume that the centres are the ones shown in Table 7, the most similar is the centre corresponding to cluster number 2.

<Insert Table 7 here>

The following step is, therefore, to recuperate all those cases with a high level of pertinence to the cluster and level of reliability from cluster nº 2 (whose centre is the most similar to the problem case). In this investigation, following the instructions of the auditors who participated in the survey, retrieving those cases with a level of pertinence to the cluster greater than 65% and whose reliability is greater than 50% has been opted for.

The retrieved cases are used in the next stage, the adaptation stage, to train the ANFIS network. In the example given, after training the network with the cases obtained in the retrieval phase, this produces the initial solution shown in Table 8.

<Insert Table 8 here>

In the phase following the ISA (Identification of the State of Activity) subsystem, the revision phase, if the internal auditor of the company which is being analysed considers that the estimation carried out is coherent, the retention phase is moved onto. Here, the initial problem is stored by the system, along with the final solution or the proposed estimation, as a new case (the addition of a new case to the case base implies the redistribution of the clusters and provokes the ANFIS network to use the new case to carry out a complete learning cycle). Otherwise, the auditor modifies the initial solution introducing a new value that, along with the problem case, will be stored as a new case. Starting with the final solution or estimation, using a rule-based system, the control risk level associated with the activity state is worked out.

Next, the second subsystem starts to work, the GR subsystem. The aim of this subsystem is to generate the recommendations that allow the improvement of the activity state. In order to be able to recommend changes in the execution of the managerial processes it is necessary to compare the current activity situation, represented by problem case + solution, with those case base cases that reflect a better managerial administration.

With this aim, only the cases most similar to the problem case will be worked with. Therefore, in the retrieval phase the same cluster used during this phase in the first subsystem is worked with (because it contains the cases most similar to the problem case). All those cases with a high level of reliability, with a solution or activity state greater in interval, by 15% to 20%, to the one generated as final solution in the first subsystem are retrieved. If enough cases are not retrieved (25 is considered to be enough) the percentages are relaxed further, increasing the percentage interval range by 5%.

Continuing with the example previously given, if we consider that Table 8 contains the final solution provided by the first subsystem, it can be observed that the obtained percentage as final solution is 58%. Therefore, the generation of recommendations subsystem's retrieval stage will consist of obtaining the c cases that present a framed solution in the 73% ...78% interval (the percentage that is obtained starting from the consideration that, given that the final solution is: 58%, the retrieved cases will have a percentage greater than 15% - 20% as a solution) from cluster n° 2. In this case we are going to assume that it is not necessary to "relax" the requested interval given that more than 25 cases are retrieved from the case base.

With this, the one that the company can improve is searched for, but always bearing in mind that the changes in the managerial sector must be carried out gradually, with small steps that imply changes that the company can adapt to progressively.

<Insert Table 9 here>

Next, in the adaptation phase, applying the Electre method, the most optimal of all the retrieved cases is chosen. Let us assume that, after applying the aforementioned method, the obtained output is the one shown in Table 9 (from here on this case will be referred to as the optimal case). The next step, for the generation of recommendations, will be to compare the optimal case with the problem case to see which tasks must be improved. For this, basing it on the level of importance ( $IR_i$ ) of the optimal case's tasks, the order of the analysis of the tasks is decided on, always starting with the ones with greater weighting or level of importance and gradually descending in importance. As we can observe, both cases do not share the same level of importance for the tasks, which implies having to decide which case is going to be taken as a guide in the generation of recommendations. Two examples are presented:

- a. Choose the problem case. This presents the values introduced by the internal auditor, valuing each task and assigning a level of importance regarding the activity group and within the company in which it is working. For this the internal auditor uses their experience as a base, which, in some cases, can be quite limited.
- b. Choose the optimal case, obtained after the applying the Electre method to the  $c$  cases retrieved from the case base, that represent the experience of a series of experts in the function within which the activity is found (in this example, experts in "Acquisition of Goods and Services").

In view of this situation, taking the weightings or levels of importance ( $IR_i$ ) of the tasks present in the optimal case as references, has been opted for. This choice is based on the fact that this case reflects the experience acquired over years by experts in the different functional areas, which implies a greater knowledge of the level of influence, of the weighting that the tasks can exert over the activity. In this way, the proposed example would start analysing the attributes corresponding to the third and fourth tasks, because of being those that present the greatest level of importance (Very High Importance), as shown in Figure 10. Specifically, the process consists of generating a recommendation that proposes the improvement of that task whose value ( $V_i$ ) in the problem case is less than the value of the same task in the case proportioned by the Electre method. In the example, the value of the third task (“Analysis of prices, discounts and credit conditions”) is 8 in the problem case and 9 in the other case, used as an aim to be achieved. Therefore, a recommendation of the “the analysis of prices, discounts and credit conditions must be improved” type will be generated.

For the fourth task the values coincide and so no recommendation would be generated.

<Insert Figure 10 here>

The process would continue with comparing the task values whose level of importance, in the case obtained after applying the Electre method, had High Importance (HI) as its value. Next, the same process with the tasks with a level of importance with Medium Importance (IM) as its value would be carried out.

## **5. Results**

A complete set of tests has been carried out over a total amount of 22 small to medium companies, for a period of 38 months. From the total number of 22, 12 were medium-sized, while ten were small sized firms, all of them pertaining to the textile sector.

Spanish auditors performed 98 surveys in order to obtain the data that would feed the process of generating the prototype cases, needed to build the system's base classes. Another 34 surveys were carried out by different experts of functional areas of firms within this sector.

In order to test the developed system various complete operation cycles of the aforementioned system were carried out. In other words, given a company, each of its activities was evaluated, a level of risk was obtained and recommendations were generated. These recommendations were communicated to the company's internal auditor and they were given a period of three months so that they could elaborate and apply an action plan, based on the previously mentioned recommendations (see Figure 11). The aim of this action plan was to reduce the number of inconsistent processes within the company. Once these three months had ended, the company was analysed again and the obtained results were compared with those obtained in the previous trimester.

<Insert Figure 11 here>

The obtained results demonstrate that the application of the recommendations generated by the system causes a positive evolution in firms. This evolution is reflected in the reduction of erroneous processes. The considered indicator to determine the positive evolution of the companies was the state of each of the analysed activities. If, after analysing one of the company's activities, it is proven that the state of the activity in question (with a value between 1 and 100) has increased with respect to the state obtained in the previous trimester, it can be said that the erroneous processes have been reduced within that activity. If this improvement is produced in the majority of the activities (above all in those that have the most relevance within the company), the company has improved its state.

With the aim of reflecting, in a more reliable way, the aptitude of the system for the problem to be resolved, the results obtained after analysing these 22 companies were compared with those of another 5 companies in which the recommendations generated by the system were not applied. In these five companies, the activities were analysed at the start of the trimester and at the end, only using the ISA (Identification of the State of the Activity) subsystem. The recommendations generated by the second subsystem were not taken into account (and therefore the recommendations in question were not applied).

In order to be able to analyse the obtained results, it must be taken into account that some of the generated recommendations implied costs that the companies could not afford or that entailed a long installation period. Therefore, a company is considered to have followed the system recommendations if they fulfil 70% or more of them.

Subsequently the obtained results are shown:

1. Of the 22 companies analysed, in those in which the recommendations generated by the system were applied, the results were (see Figure 12):
  - a) In 15 companies, the number of inconsistent processes was reduced, improving the state of activities by an average of 11.6%.
  - b) In 6 of these companies, the state of activities was improvement by an average of 0.86%. In other words, the application of the recommendations generated by the system did not have any effect on the activities of the company. After studying the possible reasons for the results, it has been concluded that the recommendations given were not completely followed, with only certain measures applied and the majority of recommendations ignored.



- c) In one company, the inconsistent processes increased, in other words, the application of recommendations generated by the system prejudiced the positive evolution of the company. Once the situation in the company had been analysed, it was concluded that there was a high level of disorganisation, without a clearly defined set of objectives. This means that any attempt to change the business organisation actually will lead to a worse situation.

In general, it could be said that these results demonstrate the suitability of the techniques used for their integration in the developed intelligent control system. The best results occurred in the companies of smaller size. This is due to the fact that these firms have a greater facility to adapt and adopt the changes suggested by the system's recommendations.

2. On the other hand, for the 5 companies in which the recommendations generated by the system were not applied, the results were as follows: the companies improved their results, though reaching an average productivity that was 5% below the same measurement for other companies that did use the system.

<Insert Figure 12 here>

## **6. Conclusions**

This article presents a neuro-symbolic system that uses two CBR systems employed as a basis for the hybridisation of a multi-criteria decision-making method, a fuzzy clustering method, a GCS network, and an ANFIS network. Therefore, the developed model combines the complementary properties of the connectionist methods with the symbolic methods of Artificial Intelligence.

The used reasoning model can be applied in situations that satisfy the following conditions:

1. Each problem can be represented in the form of a vector of quantified values.
2. The case base should be representative of the totality of the spectrum of the problem.
3. Cases must be updated periodically.
4. Enough cases should exist to train the network.

The prototype cases used for the construction of the case base are synthetic and have been created starting from the surveys completed by auditors and experts in the different functional areas.

The system is able to estimate or identify the state of the firm's activities and their associated risk. Furthermore, the system generates recommendations that will guide the internal auditor in the elaboration of action plans to improve the processes in the firm.

The estimation in the environment of firms is difficult due to the complexity and the great dynamism of this environment. However, the developed model is able to estimate the state of the firm with precision, and propose solutions that allow the improvement of the state in question.

The system will produce better results if it is provided with cases related to the sector in which it will be used. This is due to the dependence that exists between the processes in the firms and the sector in which the company is located. Future experiments will help to identify how the constructed prototype will perform in other sectors and how it will have to be modified to improve its performance.

Although the defined model has not been tested in big firms, it is believed that it could work adequately, although changes would take place more slowly than in small and

medium firms. Steps toward this direction have been taken and it is expected to be able to evaluate the system in one of the major international companies of the textile sector.

## References

- Aamodt A. and Plaza E. 1994. Case-Based Reasoning: foundational Issues, Methodological Variations, and System Approaches. AICOM. Vol. 7. N° 1, Marzo 1994.
- American Institute of Certified Public Accountants (AICPA), Statements on Auditing Standards No. 78 (SAS No. 78). 1996. Consideraciones de la Estructura del Control Interno en una Auditoría de Estados Financieros. (Amendment to SAS núm. 55), The Auditing Standards Executive Committee, New York.
- Barba-Romero, S. and Pomeral, J. 1997. Decisiones Multicriterio. Fundamentos teóricos y utilización práctica. Colección de Economía. Servicio de publicaciones Universidad de Alcalá.
- Borrajo, L. 2003. Sistema híbrido inteligente aplicado a la auditoría de los sistemas internos. Phd Thesis. Teses de doutoramento da Universidade de Vigo. Universidade de Vigo (Spain). ISBN: 84-8158-274-3. December, 2003.
- Borrajo, L., Corchado, J.M., Yáñez, J.C., Fdez-Riverola, F. and Díaz, F. 2005. Autonomous Internal Control System for Small to Medium Firms. LNCS Volume 3620 Springer-Verlag. ISSN: 0302-9743
- Corchado, J.M., Borrajo, L., Pellicer, M.A. and Yáñez, J.C. 2004. Neuro-symbolic System for Business Internal Control. LNCS Volume 3275/2004. Springer-Verlag. ISSN:0302-9743.
- Denna, E.L., Hansen, J.V. and Meservy, R. 1991. Development and application of expert systems in audit services. Transactions on Knowledge and Data Engineering.
- Fdez-Riverola, F, Corchado, J. M. and Torres, J. 2002. An Automated Hybrid CBR System for Forecasting. Advances in Case-Based Reasoning. Craw S. and Preece A. (eds.), Springer. ISBN: 3-540-44109-3.

- Gardingen, D. and Watson, I. 1998. A Web Based Case-Based Reasoning System for HVAC Sales Support. In, Applications & Innovations in Expert Systems VI. Milne, R., Macintosh, A. & Bramer, M. (Eds.), pp. 11- 23. Springer, London. ISBN 1-85233-087-2
- Hunt, J. and Miles, R. 1994. Hybrid case-based reasoning. The Knowledge Engineering Review. Vol. 9:4. pp. 383-397.
- Köhle, M. and Merkl, D. 1996. Visualizing similarities in high dimensional input spaces with a growing and splitting neural network. Proceedings of International Conference of Artificial Neural Networks, ICANN-96, (pp. 581-586), Bochum, Germany.
- Kolodner J. 1993. Case-Based Reasoning. San Mateo. CA, Morgan Kaufmann. 1993.
- Laza R., Pavón R. and Corchado J.M. 2004. A reasoning model for CBR-BDI agents using an Adaptable Fuzzy Inference System. Lecture Notes in Artificial Intelligence, Vol. 3040, pp. 96-106, Springer.
- Lenz M., Bartsch-Spörl B., Burkhard D. and Wees S. (eds.) 1998. Case-based Reasoning Technology: From Foundations to Applications, Springer Verlag, LNAI 1400.
- Mas, J. and Ramió, C. 1997. La Auditoría Operativa en la Práctica. Ed. Marcombo, Barcelona.
- Medsker L. R. 1995. Hybrid Intelligent Systems. Kluwer Academic Publishers.
- Messier, W.F. and Hansen, J.V. 1988. Inducing rules for expert systems development: an example using default and bankruptcy data. Management Science, 34, No. 12, December, 1403-15.
- Nauck D., Klawonn F. and Kruse R. 1997. Foundations of neuro-fuzzy systems. John Wiley & sons, cop.
- Riesbeck, C.K. and Schank, R.C. 1989. Inside case-based reasoning. Lawrence Erlbaum Associates. Hillsdale, NJ.

- Romero, C. 1993. Teoría de la decisión multicriterio: Conceptos, técnicas y aplicaciones. Alianza Editorial. ISBN: 84-206-8144-X
- Sánchez, A. 1995. Los Sistemas Expertos en la Contabilidad. Biblioteca Electrónica de Contabilidad, Vol. 1, N. 2.
- Warwick, K. 1995. An overview of neural networks in control applications. Chapter in Neural Networks in Robotic Control: Theory and Application. Morris, A.S. and Zalzala, A. (eds.), Ellis Horwood
- Wasserman, P. D. 1993. Advanced Methods in Neural Computing. Van Nostrand Reinhold.
- Watson I. 1997. Applying Case-Based Reasoning: Techniques for Enterprise Systems. Morgan Kaufmann.
- Watson, I. and Marir, F. 1994. Case-Based Reasoning: A Review. The Knowledge Engineering Review. Vol. 9. No. 4. pp. 355-381.

**TABLE 1.** Case structure.

PROBLEM						SOLUTION
Case number	Input vector	Function number	Activity number	Reliability	Degree of membership	Activity State

**TABLE 2.** Two passes in the hybrid learning procedure for ANFIS

	Forward pass	Backward pass
Antecedent parameters	Fixed	Gradient descent
Consequent parameters	Least-square estimator	Fixed
Signals	Node outputs	Error signals

**TABLE 3.** Summary of technologies employed in the ISA subsystem

CBR's Phases	Technology	Input	Output
Retrieve	GCS Network	Problem case	K similar cases
Reuse	ANFIS network	Problem case K similar cases	Initial solution: state of the activity
Revise	RBS	Problem case Initial solution: state of the activity	Revised solution: state of the activity Level of risk inherent to the state of the activity
Retain	GCS network	Problem case Revised solution	New Case

**TABLE 4.** Summary of technologies employed in the GR subsystem

CBR's Phases	Technology	Input	Output
Retrieve	Fuzzy Clustering	Problem case	K similar cases
Reuse	Method ELECTRE	K similar cases	Most favourable case(s)
Retain	Manual	Most favourable case(s) (used to generate recommendations)	Modified case(s)

**TABLE 5.** Example of input details for the “Selection of Suppliers” activity.

Function	Activity	Task	Importance Rate ( $IR_i$ )	Value ( $V_i$ )
Purchases	Selection of Suppliers	Provider's current capacity and potential production level	LI	1
		Study of quality of products	HI	8
		Prices, discounts and credit conditions analysis	HI	8
		Study of conditions and deadlines of deliveries	VHI	9
		Satisfaction about previous operations	HI	6
		Determination of possible legal restrictions	LI	2
		Verification of pending legal affairs	HI	5

**TABLE 6.** Assessment of the different levels of importance of the tasks

Importance Rate	Numeric Value
VHI (Very High Importance)	5
HI (High Importance)	4
AI (Average Importance)	3
LI (Low Importance)	2
VLI (Very Low Importance)	1

**TABLE 7.** Centre of each cluster for the “Selection of Suppliers” activity.

Cluster Number representing each input vector	PROBLEM								SOLUTION
	Input Vector								Activity State
1	IR <sub>i</sub>	LI	HI	VHI	VHI	HI	LI	HI	36%
	V <sub>i</sub>	1	5	6	6	3	2	3	
2	IR <sub>i</sub>	LI	VHI	VHI	VHI	HI	HI	HI	60%
	V <sub>i</sub>	3	6	7	7	5	4	5	
3	IR <sub>i</sub>	LI	HI	VHI	VHI	HI	LI	HI	90%
	V <sub>i</sub>	5	8	9	9	7	6	7	

**TABLE 8.** Initial solution obtained by the ANFIS (“Selection of Suppliers” activity).

PROBLEM								SOLUTION
Input Vector								Activity State
IR <sub>i</sub>	LI	HI	HI	VHI	HI	LI	HI	58%
V <sub>i</sub>	1	8	8	9	6	2	5	

**TABLE 9.** Obtained solution after applying the Electre method to the retrieved case from cluster n°2 (“Selection of Suppliers” activity).

PROBLEM								SOLUTION
Input Vector								Activity State
IR <sub>i</sub>	LI	HI	VHI	VHI	HI	LI	AI	75%
V <sub>i</sub>	3	8	9	9	8	5	7	

**Figure 1.** Functions, Activities and Tasks

**Figure 2.** System reasoning process

**Figure 3.** ISA Subsystem - Identification of the State of the Activity

**Figure 4.** The point mean cases in the input environment.

**Figure 5.** Example of the set of rules for calculating the risk level.

**Figure 6.** Pseudocode of the revision stage of the Recommendation Generator subsystem.

**Figure 7.** GR Subsystem - Generation of Recommendations.

**Figure 8.** Pseudocode of the retrieval phase in the Generation of Recommendations subsystem.

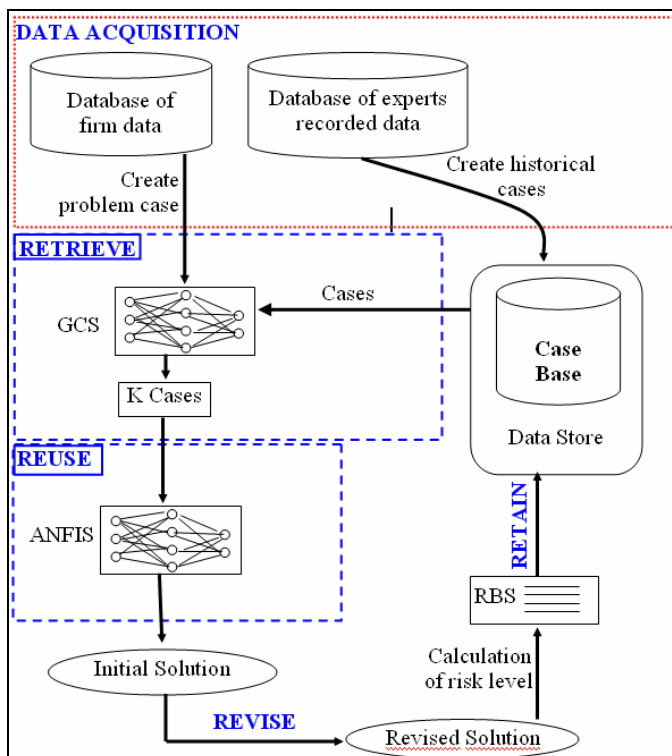
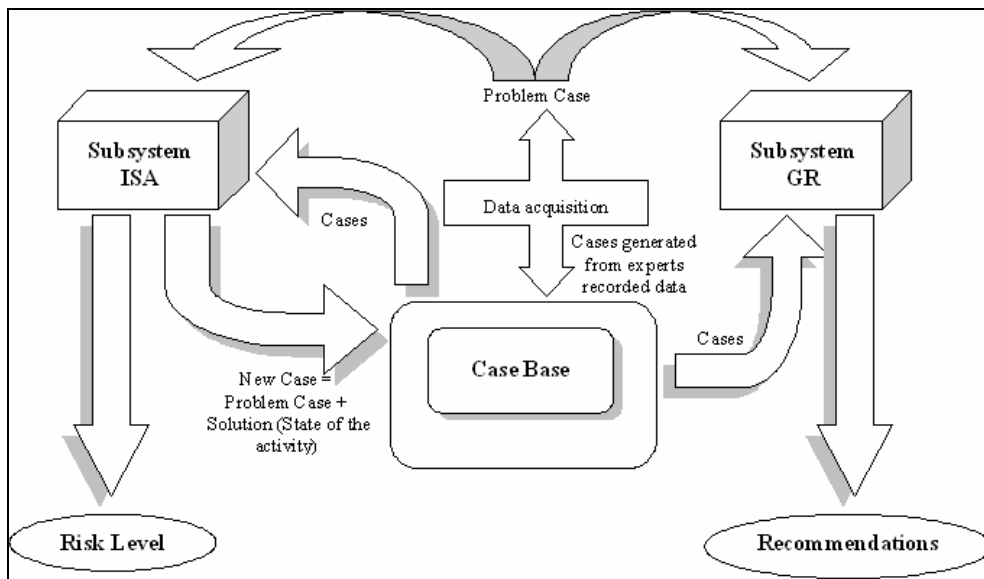
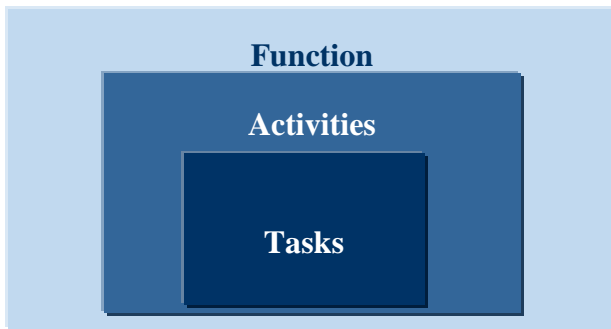
**Figure 9.** Pseudocode of the adaptation phase of the Generation of Recommendations subsystem.

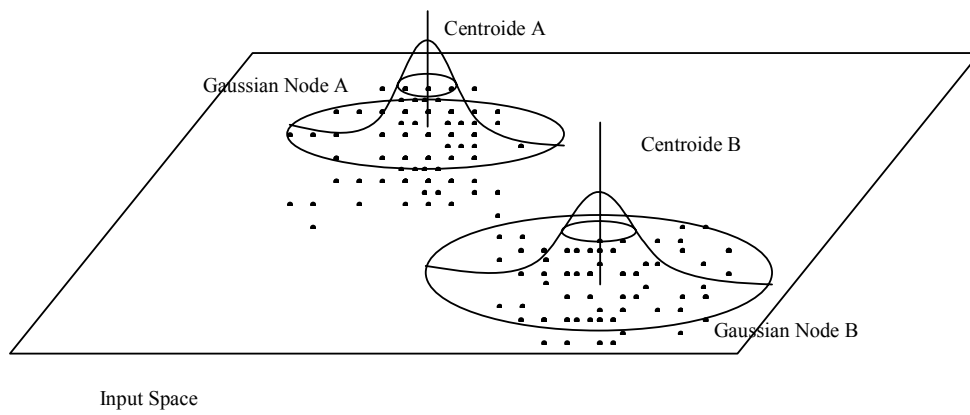
**Figure 10.** Juxtaposition of both cases to detect deviations

**Figure 11.** Operation cycle

**Figure 12.** The firm's evolution







Input Space

```

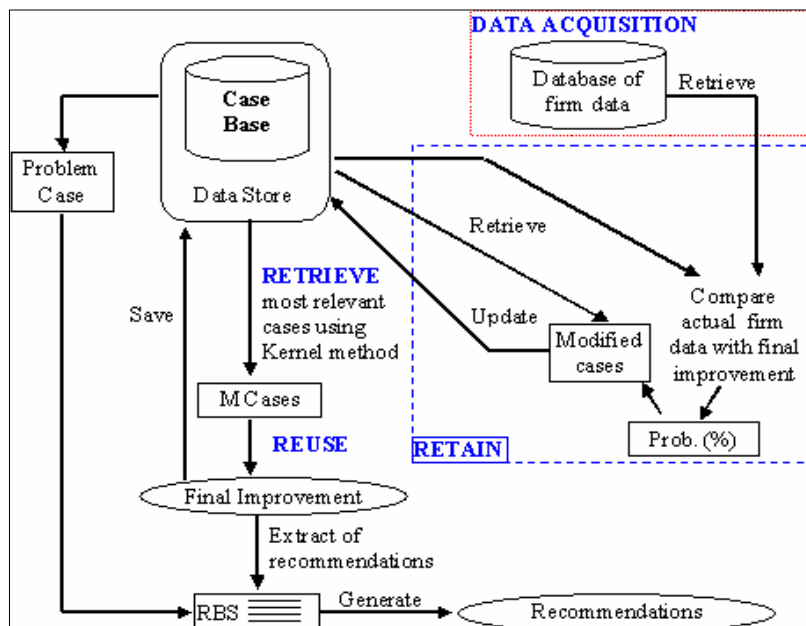
if auditors_importance = vl then
  if activity_state >= vl then
    risk_level = low
  else if activity_state < (vl-1) then
    risk_level = high
  else risk_level = average

```

```

procedure revise_EA (input:  $v_p, m_a, s_i$ ; output:  $s_f, n_r$ )
{
00   begin
01     accept  $\leftarrow$  internal_auditor_revision ( $s_i$ )
02     if accept then
03        $s_f \leftarrow s_i$ 
04     else
05        $s_f \leftarrow$  internal_auditor_proposal
06        $n_r \leftarrow$  calculate_risk_level ( $s_f, m_a$ )
07   end.
}

```



```

procedure retrieve_GR (input:  $X, v_p, s_f, m, U$ ; output:  $K$ )
{
00      begin
01           $increment \leftarrow 0$ ;
02          repeat
03              for each case  $i \in X$  do
04                   $dif \leftarrow s_i - s_f$ 
05                  if ( $dif \geq 0.15$  and  $dif \leq (0.20 + increment)$ )
06                      and ( $u_{mi} > 0.75$ )
07                      and ( $reliability(i) > 0.5$ ) then
08                           $K \leftarrow K + i$ 
09                       $increment \leftarrow increment + 0.5$ ;
08          until  $|K| > 25$ ;
09      end
}

```

```

procedure reuse_GR (input:  $K$ ; output:  $v_{el}$ )
{
00      begin
01           $(C, n) \leftarrow ELECTRE(K)$ 
02          if  $n > 1$  then
03              for each attribute  $j$  do
04                   $v_{el_j} \leftarrow \frac{\sum_{i=1}^n C_{ij}}{n}$ 
05      end
}

```

Problem Case								SOLUTION
PROBLEM								
IR <sub>i</sub>	LI	HI	HI	VHI	HI	LI	HI	58%
V <sub>i</sub>	1	8	8	9	6	2	5	

↕

↕

V <sub>i</sub>	3	8	9	9	8	5	7	75%
IR <sub>i</sub>	LI	HI	VHI	VHI	HI	LI	AI	
PROBLEM								SOLUTION
Obtained case by Electre method								

